2631 #4

OIPE JC156
APR 2 0 2001
PATENT & TRADEMARK OFFICE

RECEIVED
APR 2 6 2001
Technology Center 2600

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of | Atty. Docket No.
SYLVAIN DEVILLERS ET AL | PHF 99,613

Serial No. 09/723,426 | Group Art Unit: 2631

Filed: NOVEMBER 28, 2000

Title: METHOD FOR CODING AND DECODING MULTIMEDIA DATA

RECEIVED
APR 2 5 2001
Technology Center 2600

Commissioner for Patent
Washington, D.C. 20231

CLAIM FOR PRIORITY

Sir:

A certified copy of the European Application No. **99402972.6** filed November 29, 1999 referred to in the Declaration of the above-identified application is attached herewith.

Applicants claim the benefit of the filing date of said European application.

Respectfully submitted,

By Russell Gross
Russell Gross, Reg. 40,007
Attorney
(914) 333-9631

Enclosure

CERTIFICATE OF MAILING

It is hereby certified that this correspondence is being deposited with the United States Postal Service as first-class mail in an envelope addressed to:
COMMISSIONER OF PATENTS AND TRADEMARKS
Washington, D.C. 20231

On April 16, 2001

By Edna Chaga

F:\WPDOCS\GR\mf13grk0.ec0.doc

This Page Blank (uspto)

PHF 99.613 US

Europäisches Patentamt | European Patent Office | Office européen des brevets

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

99402972.6

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

Hatten

I.L.C. HATTEN-HECKMAN

DEN HAAG, DEN
THE HAGUE, 14/11/00
LA HAYE, LE

EPA/EPO/OEB Form 1014 - 02.91

This Page Blank (uspto)

Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

Blatt 2 der Bescheinigung
Sheet 2 of the certificate
Page 2 de l'attestation

Anmeldung Nr.:
Application no.: 99402972.6
Demande n°:

Anmeldetag:
Date of filing: 29/11/99
Date de dépôt:

Anmelder:
Applicant(s):
Demandeur(s):
Koninklijke Philips Electronics N.V.
5621 BA Eindhoven
NETHERLANDS

Bezeichnung der Erfindung:
Title of the invention:
Titre de l'invention:
"Shape and shape deformation descriptors"

In Anspruch genommene Prioriät(en) / Priority(ies) claimed / Priorité(s) revendiquée(s)

Staat:	Tag:	Aktenzeichen:
State:	Date:	File no.
Pays:	Date:	Numéro de dépôt:

Internationale Patentklassifikation:
International Patent classification:
Classification internationale des brevets:
/

Am Anmeldetag benannte Vertragstaaten:
Contracting states designated at date of filing: AT/BE/CH/CY/DE/DK/ES/FI/FR/GB/GR/IE/IT/LI/LU/MC/NL/PT/SE/TR
Etats contractants désignés lors du depôt:

Bemerkungen:
Remarks:
Remarques:

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG99/xxxx
October 1999

Source: Laboratoires d'Electronique Philips (LEP)
Status: Proposal
Title: Shape and shape deformation descriptors based on complex Fourier descriptors
Author: Sylvain Devillers, Jean-Christophe Broudin

Abstract

In this contribution, we propose two new descriptors based on complex Fourier descriptors to characterize a shape and its deformation in time. The aim of the latter is to characterize a segmented moving object as more or less rigid. For this, we use a method based on complex Fourier descriptors.

1. Introduction

Much semantic information may be extracted from the shape of an object and its deformation in time. For example, in a video-surveillance application, the rigidity of a moving region will allow to differentiate pedestrians from vehicles.
However, when vehicles are driving away from the camera, the 2D shape changes due to the perspective effect. This may be locally approximated as an affine transformation. In order to cope with this possible variation of scale or translation, the descriptor has to be invariant to basic geometrical transformations such as translation, rotation or scaling. Furthermore, we need a scalable descriptor to be able to describe the shape and its deformation with more or less precision.
To answer this need, we developed a new shape descriptor based on complex Fourier descriptors, which is invariant by translation, rotation and scaling. Then we extracted a compact shape deformation descriptor by measuring the variability of the different frequencies in time.

2. Complex Fourier Descriptors: definition and properties

2.1. Definition

Complex Fourier descriptors consist in a lossless representation of a shape contour G. A contour is defined as a set of points surrounding a surface. Depending on the sampling, points are not necessarily connex. The length of the contour is the number of points used to describe it and therefore depends on the sampling.

Complex Fourier descriptors are an equivalent frequencial description and not a parametric representation. They are defined by:

$$Z_k = \sum_{n=1}^{L} z_n \exp\left(\frac{2*i*\pi*n*k}{N}\right), \qquad 0 \le k < N$$

Equation 2-1

$z_n = x_n + iy$ stands for the coordinate of the n^{th} point of G, written as a complex number. Real part is absciss and imaginary part is ordinate.
L stands for the length of G and N for the number of frequency bins.

These descriptors have the same meaning as in signal processing:

- Low frequencies, for k around 0 and $N-1$, give a coarse idea of the shape
- High frequencies, for k around $\frac{N}{2}$) represent fine details.

This means that if two contours are very similar but for small details or for a small local part, the first coefficients will be very close, whereas the last ones will be completely different. Besides, if the shape is not rigid, the shape contour will change and so do the first coefficients. Of course the last ones will change as well, but will not be significant. Hence, first coefficients aim at clustering shape contours. Intrinsically, complex Fourier descriptors are a scalable representation of the contour.

- Z_0 stands for the continuous component (DC or Direct Current) and represents the non-normalized centroid of the contour.
- Z_1 is the radius of the circle whose surface is equivalent to that of the shape, which can be interpreted as a scale parameter.
- Z_k and Z_{N-k} , $1 < k < N$, have similar but opposite properties. k is the number of actions regularly spaced on the unity circle. k, $1 < k \le \frac{N}{2}$ represents the number of tension actions on the unity circle towards the outside, whereas k, $\frac{N}{2} + 1 < k < N$ represents the number of pressure actions towards the inside.
- The phase ϕ_k of Z_k locates the action on the circle.

2.2. Properties

2.2.1. Preliminaries

Suppose that a contour Γ_1 is translated by $\vec{T}$, rotated by ϕ and scaled by a factor λ to obtain Γ_2, such as Γ_1 and Γ_2 have the same number of points. Then there exists a simple relation between the complex Fourier descriptors Z_k^1, $0 \le k < N$ of Γ_1 and Z_k^2, $0 \le k < N$ of Γ_2:

$$Z_k^{\,2} = \vec{T} + \lambda \exp(i\phi) Z_k^{\,1}, \qquad 0 \le k < N$$

Equation 2-2

Making Z_k invariant by translation, rotation and scaling is equivalent to cancelling the effects of $\vec{T}$, ϕ and λ.

2.2.2. Translation invariance

$\vec{T}$ is a continuous component and is therefore contained in Z_0.
By not considering Z_0, the set of coefficients $\{\ Z_k, \quad 1 \le k < N\ \}$ is translation invariant.

2.2.3. Rotation and starting point invariance

By then considering the set of coefficients $\{\ abs(Z_k), \quad 1 \le k < N\ \}$ where $abs(\ )$ is the modulus of Z_k, $\{\ abs(Z_k), \quad 1 \le k < N\ \}$ is phase invariant. As both the starting point and a rotation induce a move of the phase, that is a multiplication by $\exp(i\phi)$, the descriptor is rotation and starting point invariant.

2.2.4. Scale factor invariance

We are now focusing on:

$$abs\left(Z_k^{\,2}\right) = \lambda abs\left(Z_k^{\,1}\right), \qquad 1 \le k < N$$

By finally considering the set of coefficients $\left\{\ abs\left(\frac{Z_k}{Z_1}\right), \quad 1 \le k < N\ \right\}$, the resulting descriptor is also scale invariant.

Unfortunately, λ is not known, but present in each Z_k. It is chosen to normalize by one of the descriptors. Since Z_1 is known to be a scale factor, each $abs(Z_k), \quad 1 < k < N$ will be divided by $abs(Z_1)$.

Hence, $\left\{\ abs\left(\frac{Z_k}{Z_1}\right), \quad 1 < k < N\ \right\}$ is translation, rotation and scale invariant.

2.2.5. Contour length invariance

Equation 2-1 was established for two contours of same number of points . If their number of points differ, then their frequencial description will also defer; difference of length can be interpreted as a difference of

sampling. To cancel the influence of length, one contour must be resampled to the length of the other.

By choosing for L a power of two, we can take $N = L$, which makes the description also sampling invariant. As a matter of fact, if we downsample the contour Γ from $L_1 = 2^{m_1}$ points to $L_2 = 2^{m_2}$, $m_2 < m_1$, then the first and last frequency bins of each descriptor will correspond exactly to the same frequency, because the frequency lap $\frac{N}{L}$ remains the same. Conversely for upsampling from $L_1 = 2^{m_1}$ points to $L_2 = 2^{m_2}$, $m_2 > m_1$.

2.2.6. Compaction property

$\{ Z_k, \quad 0 \le k \le N_0 \cup N - N_0 \le k < N \}$ will be a truncated list of the complete list of the N complex Fourier descriptors necessary to describe the shape losslessly. The resulting reconstructed shape will be a filtered version of the initial shape. The number N_0, $1 \le N_0 < N$ of descriptors to retain depends on the complexity of the contour. However, 50% of all coefficients are necessary to obtain a well-reconstructed contour with very few artifacts.

2.2.7. Robustness to incomplete view

Experiments have shown that the Fourier descriptors are sometimes very similar, sometimes completely different. As it both depends on the contour and the percentage of occlusion, it is safe to say it is not robust to incomplete view.

2.2.8. Scalability

As stated in 2.1, complex Fourier descriptors are intrinsically scalable: the higher the frequency, the finest the description.

3. Shape descriptor

3.1. Descriptor definition

The input data is a binary mask of an object sampled on a regular grid. The object has no holes. It is not a fractal object either. Beforehand, the contour of the object must be extracted, then resampled for its number of points to be a power of two $L_2 = 2^m$. That way, we take $N = L_2$ in the FFT.

3.2. Specifications of the proposed shape descriptor

The descriptor should not only contain the necessary information on the shape but also be a supreme

summary of the full information available at start.

We propose the following MPEG-7 descriptor.

- **Centroid** $(C_x C_y)$: coordinates of the centroid of the contour.
- **Angle** θ : angle between horizontal and main axis of the contour.
- **Size of the original contour** N : size of the contour after resampling.
- **Set of ordered Fourier coefficients** Z'_k : set of invariant Fourier coefficients.
- **Size of the Fourier coefficients set** P : size of the preceding set, $1 < P \leq N$. P is necessarily odd.
- **Scale** : scale parameter.

Here is the corresponding C structure:

```
typedef struct Shape Descriptor  {

  /* Centroid */
  long center x;
  long center y;

  /* Angle */
  float theta;

  /* Size of the original contour, after resampling (N) */
  long size of contour;

  /* Set of Fourier coefficients */
  float *Fourier Coefficients;

  /* Size of the set of Fourier coefficients (P) */
  long size Fourier Descriptors Set;

 };
```

3.3. Extraction

These are the steps which lead to a set of invariant Fourier coefficients:

- Compute the two eigenvectors and the two eigenvalues from the contour and store angle as the angle θ between the eigenvector associated with the biggest eigenvalue and the horizontal. θ is known modulo π.
- Compute the FFT on the resampled contour of size N, to obtain $\{Z_k, \quad 0 \leq k < N\}$.
  Store centroid as:

$$C_x = \frac{Re(Z_0)}{N}$$

$$C_y = \frac{\text{Im}(Z_0)}{N}$$

- Take modulus of each $\{Z_k, \quad 1 \le k < N\}$.
- Store $scale$ as:

$$scale = \frac{abs(Z_1)}{N}$$

- Divide each $\{Z_k, \quad 2 \le k < N\}$ by $abs(Z_1)$ and store as Fourier coefficients $Z'_j, \quad 1 \le j \le N$.
- Depending on the application, choose the final number P out of N Fourier coefficients to keep.
- Store Fourier coefficients $abs\left(\frac{Z_k}{Z_1}\right)$ in the following order:

$$Z'_1 \dots Z'_P = abs\left(\frac{Z_2}{Z_1}\right) abs\left(\frac{Z_{N-1}}{Z_1}\right) \dots abs\left(\frac{Z_{\frac{P}{2}}}{Z_1}\right) abs\left(\frac{Z_{N-\frac{P}{2}+1}}{Z_1}\right)$$

3.4. Matching

Given two sets of Fourier descriptors Θ_1 and Θ_2, we want to compare their similarity. We will not take into account the position nor the angle, which do not characterize the shape itself and can be treated separately. If the two sets are of different sizes, P_1 and P_2 respectively, and for instance $P_1 < P_2$, then we must compare the first P_1 Fourier coefficients of the two sets.

Considering that for one set, values $f_{\Theta_i}(k)$ at each frequency bin of order k are of different order of magnitude, it is relevant, for each frequency bin, to normalize the difference of values between the two sets by the magnitude at the current frequency bin. To harmonize the difference of magnitude between frequencies, it has been chosen to sum relative errors between corresponding frequency bin values $f_{\Theta_i}(k)$ of the two descriptors.

Finally, we should consider that the coarse structure (low frequencies) prevails over fine details (high frequencies), and therefore introduce a weighting function $\omega(k)$ which privileges low frequency range at the expense of high frequency range. Therefire, it sets the influence of details in the final result.

Δ will denote the dissimilarity function and Λ the corresponding similarity function. Return values are between 0 and 1.

$$\Delta(\Theta\ ,\Theta_2)=\frac{\sum_{k=1}^{P}2\omega(k)\mathrm{E}(Z'_{\Theta 1}(k),Z'_{\Theta 2}(k))}{\Omega}$$

$$P=\min(P_1,P_2)$$

$$\omega(k)=\frac{1}{1+k^2}$$

$$\begin{aligned}\mathrm{E}(x,y) &= \frac{x-y}{x} \qquad si \qquad x>y\\ &= \frac{y-x}{y} \qquad si \qquad y>x\end{aligned}$$

$$\Omega=\sum_{k=1}^{P}2\omega(k)$$

$$Sim=1-\Delta(\Theta\ ,\Theta_2)$$

4. Shape deformation descriptor

4.1. Descriptor definition

The input data is a segmented video sequence of an unique object, that is a sequence of binary masks. The shape descriptor of the contour at each frame will be computed stored for processing, as described in 3.3

This descriptor is based upon the shape descriptor exposed above.

- **Normalized deviation of the scale :** normalized deviation of the scale parameter over the video sequence.
- **Maximal size of the original contours** N_{max} **:** the maximal size of the original contour sizes N over the video sequence. N is an item of the shape descriptor.
- **Normalized deviations of each Fourier coefficient** $\sigma_{z_k'}$ **:** normalized deviations of each Fourier coefficient over the video sequence.
- **Size of the set of normalized deviations of each Fourier coefficient** M **:** size of the preceding set.

Here is the corresponding C structure:

```
typedef struct ShapeDeformationDescriptor {

  /* Normalized deviation of scale */
  float Deviation of scale;

  /* Maximal size of the original contours in the video sequence (N max)
*/
  long Maximal Size of Original contours;
```

```
/* Normalized deviation on Fourier coefficients */
float *Deviaton of Fourier Coefficients;

/* Size of the set of normalized deviations of Fourier coefficients */
 long Size of Fourier Coefficients Set;

};
```

4.2. Extraction

Deviation of the scale factor and of each Fourier coefficient over the video sequence is calculated by using the standard deviation. Dividing by the mean provides a normalization of the deviation.

The size of the set of Fourier coefficients may vary along the video sequence, but as the frequency lap remains the same, k^{th} Fourier coefficient $Z_k'^{i}$ of i^{th} frame will be averaged with the k^{th} Fourier coefficient $Z_k'^{j}$ of j^{th} frame.

- Calculate the mean of $scale$ over the video sequence,
- Calculate the mean of each Fourier coefficient Z_k' over the video sequence,
- Calculate the standard deviation of $scale$ over the video sequence,
- Calculate the standard deviation of each Fourier coefficient Z_k' over the video sequence,
- Divide the standard deviation of $scale$ by its mean and store as σ_{scale},
- Divide each Z_k' by its mean and store as $\sigma_{Z_k'}$

4.3. Matching

Although a matching function is not relevant for our shape deformation descriptor, because shape deformation descriptors are not intended to be compared, we provide one anyway. The following function quantifies the similarity between two shape deformation descriptors θ_1 and θ_2.

The number of normalized deviations of Fourier coefficients implied in the calculation depends on the sizes M_1 and M_2 of the two sets of normalized deviations to compare.

A weighting function $w(k)$ privileges low frequency range at the expense of high frequency range, to set the influence of details in the final result.

Δ will denote the dissimilarity function and Λ the corresponding similarity function.

$$\Delta(\Theta\ ,\Theta_2)=\frac{\sum_{k=1}^{M}2\omega(k)\mathrm{E}\left(\sigma_{Z'_{1,k}},\sigma_{Z'_{2,k}}\right)}{\Omega}$$

$$M=\min(M_1,M_2)$$

$$\omega(k)=\frac{1}{1+k^2}$$

$$\mathrm{E}(x,y)=\frac{x-y}{x} \quad si \quad x>y$$

$$=\frac{y-x}{y} \quad si \quad y>x$$

$$\Omega=\sum_{k=1}^{M}2\omega(k)$$

$$Sim=1-\Delta(\Theta\ ,\Theta_2)$$

5. Conclusion

Present document introduced two descriptors, a shape descriptor for still images in chapter 3 and a shape deformation descriptor for objects in video sequences in chapter 4 .

The shape descriptor is based on complex Fourier descriptors whose theory has been explained in 2.1 . It gives a frequencial description of the contour of the object.

First results show that the shape descriptor is both robust and discriminating. It is invariant by translation, rotation and scaling and also scalable. It handles resampling. Tests have even proved that downsampling increase matching scores. The dedicated matching function allows to set the degree of similarity between two objects, as explained in 3.4.

This shape descriptor is used as a basis to characterize shape deformation in a video sequence and thus to define a percentage of variation of each Fourier coefficient. That is possible because of the meaningful interpretation of its frequencial description.

First results presented indicate that it is possible to evaluate how much a shape can be deformed, by looking at . Its normalized deviation appears to quantify the degree of deformation. Its value is the deformation rate. Even if it is not designed for, this descriptor can be considered a signature of the shape deformation and be used in a query search to match objects that get out of shape in the same way.

CLAIMS :

1. A descriptor for the representation of a shape of a moving region, said descriptor giving a way to characterize a segmented moving object as more or less rigid on the basis of input data available in the form of a binary mask of said object, sampled on a regular grid.

2. A descriptor scheme using a shape descriptor according to claim 1.

3. A video-surveillance device including a description scheme according to claim 2.

4. A descriptor for the representation of a shape of a moving region, said descriptor giving a way to characterize the deformation of a segmented moving object on the basis of input data available in the form of a sequence of binary masks corresponding to a segmented video sequence corresponding to said object.

5. A description scheme using a shape deformation scheme according to claim 4.

6. A video-surveillance device including a description scheme according to claim 5.

Abstract

The invention relates to a representation mode of the shape of a region, provided for characterizing a segmented moving object, and to a description scheme using such a shape descriptor.

The invention also relates to a representation mode of the deformation of such a shape, provided for characterizing the deformation of an unique moving object within a segmented video sequence.

The invention finally relates to apparatuses such as video-surveillance devices, including description schemes that use such shape and shape deformation descriptors. The main application of the invention is the future MPEG-7 standard.